



Universidade Presbiteriana Mackenzie

TECNOLOGIA EM BANCO DE DADOS - 2025

Análise de Dados sobre Mortes em Excesso Associadas à COVID-19: Estudo Aplicado de Ciência de Dados

DISCIPLINA: PROJETO APLICADO I - BANCO DE DADOS - 2025.2

GRUPO 11:

DANIEL SILVA SOUZA FILHO

<https://github.com/danielfilho-school/Grupo11-PA1>

2025

SUMÁRIO

Apresentação do Projeto.....	3
Apresentação da Organização.....	3
Área do Problema.....	3
Descrição do Problema / Gap.....	4
Proposta Analítica.....	4
Dados Disponíveis.....	5
Análise Exploratória.....	5
Resultados Pretendidos.....	6
Referências:.....	8

Análise de Dados sobre Mortes em Excesso Associadas à COVID-19: Estudo Aplicado de Ciência de Dados

DANIEL SILVA SOUZA FILHO

(Os demais integrantes, Caio e Naiara, nunca responderam às mensagens, e não houve qualquer comunicação da parte deles – conforme já detalhei previamente ao professor por e-mail)

Apresentação do Projeto

Este trabalho tem como objetivo analisar o impacto global da pandemia de COVID-19 por meio do dataset de Global Excess Deaths Associated with COVID-19 disponibilizado pela Organização Mundial da Saúde (OMS), desenvolvendo uma narrativa analítica que traduza dados complexos em conhecimento acessível.

Apresentação da Organização

Organização Mundial da Saúde (OMS) – A OMS é a principal agência internacional de saúde pública, responsável por monitorar, padronizar e divulgar informações epidemiológicas e sanitárias para mais de 190 países. Seu posicionamento é de referência científica global, com forte atuação em vigilância epidemiológica, modelagem de dados e políticas públicas.

O estudo analisado faz parte das iniciativas de Data Science do WHO Data Hub, plataforma oficial de dados abertos da instituição.

Área do Problema

Durante a pandemia de COVID-19, muitos países apresentaram:

- Subnotificação de óbitos;
- Falhas de registro civil;
- Diferenças metodológicas nos critérios de confirmação;

- Sobrecarga hospitalar que aumentou mortes indiretas.

Isso torna insuficiente confiar apenas em “mortes oficialmente reportadas por COVID-19”.

Descrição do Problema / Gap

O problema central é: *As mortes oficialmente reportadas não refletem o impacto real da pandemia, criando um gap entre a mortalidade registrada e a mortalidade real.*

Esse gap impede:

- comparações entre países,
- análises temporais confiáveis,
- avaliação adequada do impacto sanitário,
- estimativas de políticas públicas eficientes.

A métrica de mortes em excesso surge como solução para preencher essa lacuna.

Proposta Analítica

A proposta consiste em:

1. Consolidar o dataset da OMS (2020–2021).
2. Analisar padrões temporais e geográficos do excesso de mortalidade.
3. Comparar países e regiões segundo:
 - Excesso absoluto,
 - P-Score,

- Taxas por 100 mil habitantes,
- Incerteza estatística.

4. Elaborar um Storytelling Visual com:

- Séries temporais,
- Rankings,
- Histogramas de distribuição,
- Perfis por país,
- Interpretação contextual.

Dados Disponíveis

O dataset contém:

- País/Região
- Data (ano/mês)
- Mortes observadas
- Mortes esperadas
- Mortes em excesso
- Excess deaths per 100k
- Excess P-Score (%)
- Intervalos de incerteza (lower / upper)

Tipo: CSV/XLSX

Origem: WHO Data Hub

Sensibilidade: Baixa (dados públicos agregados)

Proprietário: OMS

Limitações: ausência de dados completos em alguns países

Análise Exploratória

A EDA revisada no Python mostrou que:

1. Evolução temporal global:

- Houve aumento significativo de mortalidade excessiva em 2021.
- Vários picos estão alinhados com variantes de maior transmissibilidade.

2. Distribuição global:

- A distribuição apresenta cauda longa à direita: poucos países concentraram grandes quantidades absolutas de excesso.

3. Ranking Top/Bottom:

- Top: Índia, EUA, Rússia – valores absolutos maiores.
- Bottom: países com baixa população ou inconsistência de dados.

4. Padrões por país:

- Alguns países apresentam vários picos (ex.: Brasil).
- Outros apresentam eventos mais concentrados (ex.: China).

5. Cobertura:

- Países com baixa qualidade de registro civil apresentam maior quantidade de NAs e intervalos de incerteza amplos.

Resultados Pretendidos

- Criar uma narrativa que explique o impacto real da pandemia.
- Apresentar padrões globais e regionais de mortalidade.
- Identificar possíveis causas de diferenças entre os países.

- Utilizar visualizações claras que se conectem a eventos reais.
- Contribuir para o entendimento público e acadêmico do fenômeno.

Referências:

WORLD HEALTH ORGANIZATION. Global excess deaths associated with COVID-19 (modelled estimates). WHO Data. Disponível em: <https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates>. Acesso em: [15/out/2025].

GOLDSCHMIDT, Ronaldo; BEZERRA, Eduardo. Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. 2. ed. Rio de Janeiro: Elsevier, 2015.

CAPPRA, Ricardo. Rastreável: redes, vírus, dados e tecnologias para proteger e vigiar a sociedade. São Paulo: Actual, 2021.